Estimating the Black-White Parity in Returns to Education

Jack Rosetti *

March 10, 2022

Abstract

This paper seeks to find precise estimates for different returns to education using an extended Mincer equation. I group workers by their commuting zones, which is the best way to measure local economies to include both metropolitan and agricultural centers.

Introduction

Do all races face the same returns to their education? This question is particularly important for economists who wish to estimate returns on investment in human capital. The seminal equation to estimate returns to education comes from Mincer (1974) who posits that the logarithm of wages is equal to one's education level (in years), experience (also in years) and a squared experience squared term to capture potential experience. There are some serious issues with the equation Mincer (1974) proposes, despite it being a good start. Firstly, it assumes that there is no difference in returns to education between racial groups. Additionally, it assumes that geography and local labor markets do not factor into one's return on investment. Therefore, this paper seeks to estimate different returns to educations by taking geography and race into account.

*Email: jrosetti@ucsd.edu

1 Theoretical Foundations

There is good reason to suspect that not all races earn the same returns to each year of investment. Firstly, it elides the fact that labor market discrimination exists on a different scale for different demographic groups. Lang and Manove (2011) for example finds evidence that even after controlling for school quality, blacks and whites earn different amounts. This is due to, in their words, ”the operation of the labor market” Lang and Manove (2011). Indeed, they find significant evidence of labor market discrimination. This motivates our use of a more robust Mincer estimation equation: we would like to capture heterogeneity across different races and geographies.

Additionally, there is the problem of purchasing power parity across geographies. Put differently, some counties have higher price levels than others, thereby necessitating higher wages. This is not reflected in the original Mincer formulation. Additionally, different local economies have different compositions in terms of high- and low-skill job positions. The Rust Belt, for example, famously has a higher proportion of low-skill jobs than a place like Southern California, leading to a different wage level based on productivity.

2 Data

Data for this study comes from the American Community Survey (ACS) and Autor and Dorn (2013). The ACS provides representative samples with person weights to adjust standard errors. Data from Dorn allows us to aggregate counties into commuter zones. This sample comes from the 2019 survey year of the ACS. Data from the ACS was chosen because of its reliability and large sample size. Despite us only using the sample of one year, we still have enough samples for each commuter zone and state to make all of the normality assumptions.

After removing people who do not work at least 50 weeks in the year for at least 35 hours per week (i.e., all full-time workers), we are left with a sample size of 657,165. Our sample contains information about hourly earnings for full-time workers across different races and commuter zones.

To calculate experience, we subtract years of school from age, and subtract five from that. This gives us a rough approximate for years of industry experience.

Summary statistics are shown in Table One. I also provide two non-paramtetric figures showing the national relationship between race and

Variables	Summary				
	Obs	Mean	Std. Dev.	Min	Max
Years of education	657,165	14.43	2.99	0	20
Years of experience	657,165	24.51	13.59	-4	75
Experience Squared	657,165	785.69	721.28	0	5625
Sex	657,165	1.43	.5	1	2
Age	657,165	43.94	13.19	16	80
Usual Hours Worked per Week	657,165	43.92	8.11	35	99
(Log) Wage	657,165	3.13	.74	0	5.97
Black	657,165	.09	.28	0	1
Other race	657,165	.13	.34	0	1

Table 1: Summary statistics for the 2019 ACS sample.

earnings in Figure One and Figure Two.

3 Empirical Results

We set our extended Mincer equation as follows:

$$lnwage = \alpha + \beta_1 edyears + \beta_2 experience + \beta_3 experiencesq + \beta_4 Black + \beta_5 female + \Gamma + \tau + \epsilon \quad (1)$$

Where Γ_i is birth fixed effects and τ is time fixed effects. We estimate multiple variations of this equation to see how much the point estimates change. The other coefficients are estimated parametrically using OLS and clustering standard errors at the commuter zone level.

We then estimate our specified regression. Table Two shows the output of the basic OLS estimations of the Mincer Equation and the Extended Mincer Equation. We want to get a rough idea of the estimates before and after adding more controls to see how much each point estimate changes.

Figure 1: Nonparametric distribution of male and female earnings by race.

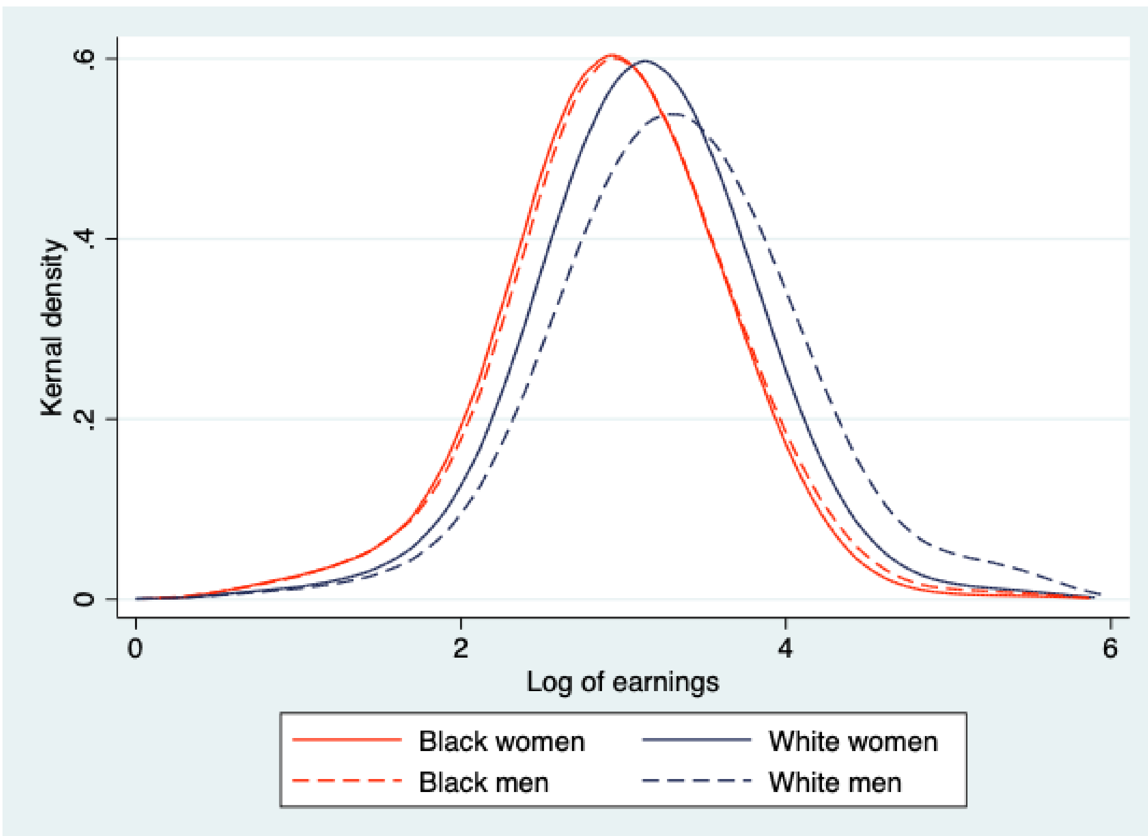


Figure 2: Nonparametric estimates of female and male earnings by race with age.

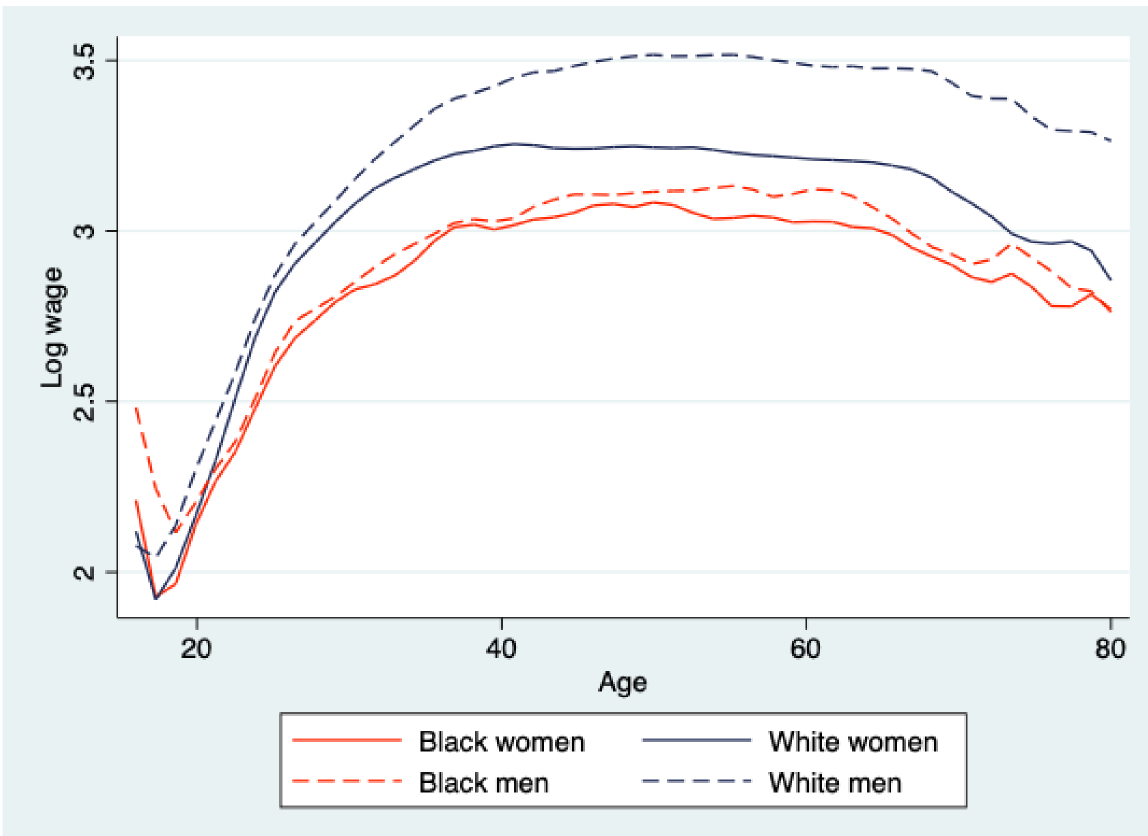
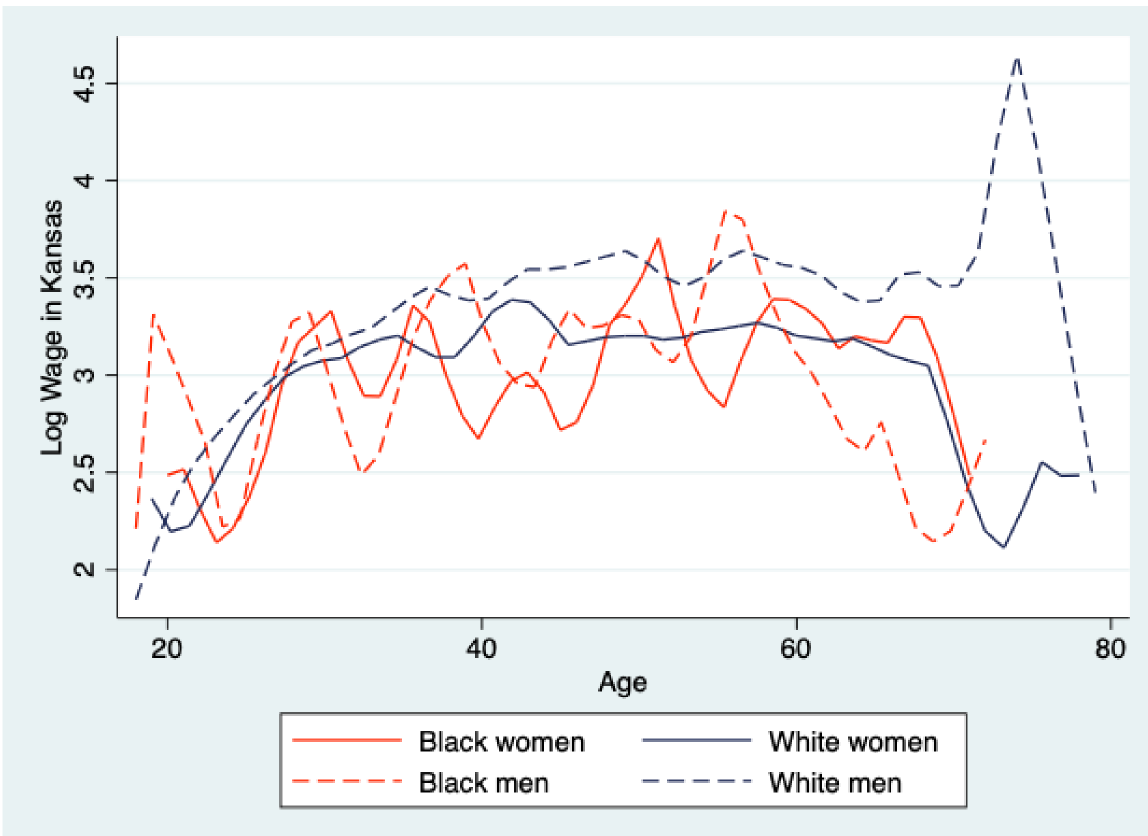


Figure 3: Nonparametric estimates of female and male earnings by race with age in Kansas.



As we see from Table 2, there is an earnings premium for white Americans. They tend to earn .223 log points more than black Americans. Additionally, we see that women face a similar earnings penalty of approximately .2 log points. All of these results are significant at the one-percent level.

Table 3 provides the output for fixed effects specifications. Individuals are grouped by commuter zones and the decade they were born. The time fixed effects are supposed to control for variation across different decades, as year-by-year earnings might not reveal as much heterogeneity. Our point estimates are slightly higher for Blacks in the fixed effects specification: they earn on average -.241 log points less than their white counterparts.

Table 4 provides an OLS estimation using interaction terms, which allows us to specify the potential penalty in returns to education for Blacks. Our estimation shows that on average, Black Americans face a penalty of .015 log points, does suggest a penalty to earnings. However, from this estimate, it is somewhat hard to measure the economic significance of the penalty, considering we are measuring log points. We propose using the delta method to examine the ratio of Black-White returns to education.

Table 5 shows the estimate for the nonlinear combination. We get that on average, the coefficient on the interaction term for Black returns to education is .135 percentage points below the coefficient on white returns to education. This result is highly economic and statistically significant, as it reveals a steep parity in wages for Black Americans entering the labor market with higher levels of education.

Our main advantage of using commuter zone data is that it provides a more granular look at earnings differentials. When we test for the greatest earnings penalties by race, we find that, on average, Kansas has the highest penalty on returns to education, as Black Americans in Kansas see an earnings penalty of -.04 log points as seen in Table 6. We plot our results in Figure 3, showing the volatility of the different returns to education. To estimate the economic significance of this gap, we estimate the ratio of earnings between Black Americans and whites. In Table 7, we get a huge earnings penalty of .58, suggesting harsh labor market discrimination.

4 Conclusion

This paper has developed a more robust estimation equation for the Mincer equation. We find significant evidence of an earnings penalty for both Black Americans and women in the labor market. We exploit interaction terms to test the economic and statistical significance of this labor market outcome;

we find that on average Blacks earn .865 percent of what whites earn for each additional year of education. More extreme cases like Kansas reveal higher earnings penalties for education.

Future work should consider larger datasets. It would also be interesting to examine wage velocity for differing groups. In other words, do Black Americans always get out-earned, or is there some ”catching up” that occurs.

It is also important to consider that these results do not reveal anything about the causal effect of education on earnings. However, these results warrant consideration for future policy work aimed at addressing earnings gaps.

References

Autor, D. H. and D. Dorn (2013, August). The growth of low-skill service jobs and the polarization of the us labor market. *American Economic Review 103*(5), 1553–97.

Lang, K. and M. Manove (2011). Education and labor market discrimination. *American Economic Review 101*(4), 1467–96.

Mincer, J. (1974). Schooling, experience, and earnings. human behavior & social institutions no. 2.

5 Appendix

Table 2: Estimations of the racial parities in returns to education

	OLS	Race Covariates	Race + Sex Covariates
Years of Education	0.109*	0.108*	0.110*
	(0.000340)	(0.000338)	(0.000339)
Experience	0.0410*	0.0413*	0.0405*
	(0.000246)	(0.000245)	(0.000243)
Experience Squared	-0.000565*	-0.000571*	-0.000557*
	(0.00000489)	(0.00000487)	(0.00000484)
Black		-0.245*	-0.221*
		(0.00258)	(0.00259)
Other		-0.0257*	-0.0218*
		(0.00225)	(0.00222)
Female			-0.213*
			(0.00158)
Observations	657165	657165	657165
Adjusted R^2	0.236	0.245	0.265

Standard errors in parentheses

These regressions are initial Mincer Equations which estimates returns to education.

Our first reduced form equation seems to do a good job estimating the point estimates on years of education and the two experience variables. However, once we include race and sex covariates, we find that on average, women and Black Americans face a steeper earnings penalty. All point estimates are statistically significant.

$^{+}$ $p < 0.10$, * $p < 0.05$

Table 3: Fixed effects estimation of earnings penalties

	Time and Commuter Zone Fixed Effects
Years of Education	0.110*
	(0.00125)
Experience	0.0249*
	(0.00152)
Experience Squared	-0.000184*
	(0.0000338)
Black	-0.241*
	(0.00599)
Other	-0.103*
	(0.0104)
Female	-0.213*
	(0.00579)
Observations	657165
Adjusted R^2	0.303

Standard errors clustered by commuter zone.
This estimation equation groups observations by commuter zones and decade of birth. This serves two purposes: we get much more precise estimates for wage differentials due to controlling for finer geographic variation. Since commuter zones are defined as local labor markets, we estimate outcomes by local labor markets. Additionally, controlling for birth cohort by decade allows us to control for heterogeneous effects across time, since different birth decades face different labor market conditions. Additionally, there is not a lot of year-by-year variation; decades provide more accurate groupings. We see again our estimates do not change by much, which points to our initial OLS being accurate. Our fixed effects model suggests a much larger earnings penalty for Black Americans, at around 24%.
$^{+}$ $p < 0.10$, * $p < 0.05$

Table 4: Fixed Effects with Interaction Terms

	Log of Wage
Years of Education	0.111*
	(0.00265)
Experience	0.0409*
	(0.000905)
Experience Squared	-0.000563*
	(0.0000173)
Black	-0.0277
	(0.0448)
Other	-0.0209
	(0.0207)
Female	-0.226*
	(0.00688)
Female × Black	0.133*
	(0.00763)
Years of Education × Black	-0.0150*
	(0.00301)
Experience × Black	-0.00404*
	(0.00109)
Experience Sq. × Black	0.0000605*
	(0.0000229)
Observations	657165
Adjusted R^2	0.266

Standard errors are clustered by commuter zone.
This table shows the fixed effects estimates with interaction terms to capture
differential labor market outcomes for Black Americans. Of interest to us is the coefficient of
Years of Education × Black considering the design of our study. We indeed find an earnings
penalty of -.0150 log points. However, it is hard to tell if it is economically significant. We develop
a formal test using the Delta Method to find the ratio of Black returns to
education to white returns to education.
change in this estimation equation.
$^{+}$ $p < 0.10$, * $p < 0.05$

Ratios	Non-linear Combination			
	Coefficient	Std. Error	z	$P > \|z\|$
(Edyrs × Black) + Edyrs/Edyrs	.865	.025	34.55	0.00

Table 5: This table provides a nonlinear combination of the coefficients for years of education and the interaction term between Black and years of education using the delta method. This result provides significant evidence of a gap between the return on education for Blacks compared to whites. Particularly, there seems to be a .135 percent earnings penalty for Blacks in the labor market.

Table 6: Estimating the earnings penalty in Kansas

	Log Wage
Years of Education	0.118*
	(0.00343)
Experience	0.0206*
	(0.0000402)
Experience Sq.	-0.0000742
	(0.0000263)
Black	0.469
	(0.0895)
Other	-0.0408
	(0.0156)
Female	-0.271^{+}
	(0.0390)
Female × Black	0.209
	(0.0681)
Years of Education× Black	-0.0493*
	(0.000404)
Experience × Black	0.00275
	(0.00205)
Experience Sq. × Black	-0.000109
	(0.0000178)
Observations	2578
Adjusted R^2	0.283

Standard errors are clustered by commuter zone.
This table shows the fixed effects estimates with interaction terms to capture differential labor market outcomes for Black Americans in Kansas. Of interest to us is the coefficient of Years of Education × Black. We indeed find an earnings penalty of -.0493 log points. However, it is hard to tell if it is economically significant. We develop a formal test using the Delta Method to find the ratio of Black returns to education to white returns to education.
$^{+}$ $p < 0.10$, * $p < 0.05$

	Kansas Non-linear Combination			
Ratios	Coefficient	Std. Error	z	$P > \|z\|$
(Edyrs $\times$ Black) + Edyrs/Edyrs	.5810085	.0156449	37.14	0.00

Table 7: This table provides a nonlinear combination of the coefficients for years of education and the interaction term between Black and years of education using the delta method. This result provides significant evidence of a gap between the return on education for Blacks compared to whites in Kansas. Particularly, there seems to be a .419 percent earnings penalty for Blacks in the labor market.